

Recursion-and-Fuzziness Reinforced Online Sparse Streaming Feature Selection

Ruiyang Xu, *Graduate Student Member, IEEE*, Di Wu , *Member, IEEE*, and Xin Luo , *Fellow, IEEE*

Abstract—Online streaming feature selection (OSFS) is a critical technique for addressing high-dimensional streaming data in various real applications. The challenge of OSFS arises from missing entries due to various reasons such as equipment failure or human mistakes. Online sparse streaming feature selection (OS²FS) is a feasible approach to this challenge by pre-estimating the missing data before feature selection based on latent factor analysis (LFA). However, such an approach separates the processes of the LFA-based estimates and the down-streaming feature selection, which cannot represent the uncertain relationships between the sparse features and the labels, thereby leading to accuracy losses. To address this critical issue, this article proposes a recursion-and-fuzziness reinforced online sparse streaming feature selection (RF-OS²FS) model, which consists of two-fold ideas: first, connecting the LFA-based feature estimation process and the consequent feature selection process via a recursively completion sampling strategy, therefore enabling the information feedback loop from the feature selection validation results to the missing feature estimation, and second, adopting the three-way decisions strategy to establish the fuzzy feature selection for representing the uncertain relationships between the sparse features and the labels. Experimental results on ten real-world benchmark datasets demonstrate that the proposed RF-OS²FS model significantly out-performs existing state-of-the-arts, both OSFS and OS²FS models in terms of accuracy when performing sparse streaming feature selection for down-streaming classification tasks.

Index Terms—Deep learning, fuzziness, latent factor analysis (LFA), missing data, online feature selection, streaming data, three-way decision (3WD), uncertainty.

NOMENCLATURE

t	Timestamp $t \in \{1, 2, \dots, T\}$.
F	Streaming features, $F = \{F_1, F_2, \dots, F_T\}$.

Received 22 March 2025; accepted 1 May 2025. Date of publication 12 May 2025; date of current version 6 August 2025. This work was supported in part by the National Key Research and Development Program of China under Grant 2024YFF0908200, in part by the National Natural Science Foundation of China under Grant 62176070 and Grant 62272078, in part by the New Chongqing Youth Innovation Talent Project and under Grant CSTB2024NSCQ-QCXMLX0035, in part by the Chongqing Technical Innovation and Application Development Special Project under Grant CSTB2023TIAD-KPX0037, and in part by the Chongqing Natural Science Foundation under Grant CSTB2023NSCQLZX0069. Recommended by Associate Editor W. Ding. (*Corresponding author: Xin Luo.*)

Ruiyang Xu is with the School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China, and also with the Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 400714, China (e-mail: d220201035@stu.cqupt.edu.cn).

Di Wu and Xin Luo are with the College of Computer and Information Science, Southwest University, Chongqing 400715, China (e-mail: wudi.cigit@gmail.com; luoxin@swu.edu.cn).

Digital Object Identifier 10.1109/TFUZZ.2025.3569272

1941-0034 © 2025 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See <https://www.ieee.org/publications/rights/index.html> for more information.

F_t	t th column feature vector of F .
$f_{m,t}$	m th row sample vector of F_t , $m \in \{1, 2, \dots, M\}$.
F'_t	t th column sparse streaming feature vector of F' .
$f'_{m,t}$	m th row sample vector of F'_t .
Λ_t	Known values of F'_t at time stamp t .
Ω_t	Enhanced values of F'_t at time stamp t .
H	Buffer's column size.
θ	Parameter of enhanced data.
ψ	Correlation enhancement function's parameter.
$u_{m,k}^d$	On d th layer, the m th row, and k th column vector of U_d .
$v_{j,k}^d$	On d th layer, The j th row and k th column vector of V_d .
$\hat{B}_{t,d}$	Completed sparse streaming features matrix.
$\hat{F}'_{t,d}$	t th column vector of the $\hat{B}_{t,d}$.

I. INTRODUCTION

WITH the advent of the era of big data, high-dimensional data have become more complex and dynamic [1], [2]. By removing irrelevance and redundancy features, feature selection can gain significant benefits from high-dimensional data [3], [4], [5]. However, features flow individually over time, and the feature space may be generated dynamically in practice [6]. Traditional feature selection models always require the feature space in advance, which makes it difficult to deal with streaming features [7]. Motivated by this, online streaming feature selection (OSFS) is designed to deal with infinite streaming features in an online manner [8]. All the above studies focus only on the completed high-dimensional streaming data and ignore missing entries.

In the actual application, the high-dimensional data often has massive missing entries [9], [10]. For example, Douban Matrix, the largest online database of books, movies, and music in China, has 58 541 items and 129 490 users but only contains 16 830 839 known data items, with a data density of only 0.22% [11]. Therefore, selecting features from sparse high-dimensional streaming data is a highly research hotspot. Considering the incomplete streaming features scenario, Wu et al. [12] proposed online sparse streaming features selection (OS²FS), named LOSSA, which utilizes known values to pre-estimate the missing entries based on the latent factor analysis (LFA) model before selecting features. However, LFA-based estimates and the consequent feature selection are independent, making a single-LFA-based model unable to gain benefits from the selected features.

Recently, deep forest utilizes the core idea of a multigrained cascade forest to establish an association between its two steps

and construct the multilayered structure, i.e., cyclically connecting multigrained scanning and cascade forest. Each layer receives the feature information processed by the previous layer and outputs it to the next stage, which can enhance the representation ability [13]. Notably, embedding the LFA model into a multilevel structure can enhance data density, thus recursively improving the effect of the LFA model to pre-estimate missing data [14]. Consequently, deep forest can provide a multilevel associated structure for OS²FS. Motivated by this, can we establish the association between the two steps of OS²FS recursively?

As we all know, the error between the completed data and actual data influences the effect of the recursively-reinforced OS²FS model. Although the recursively-reinforced structure can reduce this error layer-by-layer, it is difficult to ensure that there is no error. These inevitable errors cause the uncertain relationships between the sparse features and the labels [15]. Because of this uncertain relationship, it is impossible to determine the boundary of feature correlation accurately, i.e., an unsharp and gradually changing boundary, which leads to fuzziness [16]. However, existing OS²FS methods cannot effectively deal with such fuzziness that leads to accuracy loss. The three-way decision (3WD) is a practical, theoretical framework to address various uncertain problems by thinking in “threes.” The decision rules of 3WD theory contain acceptance, delayed, and rejection decision, which divide the whole universe into positive region, boundary region, and negative region, respectively. And then apply different strategies to these three regions [17], [18]. Inspired by this, this article adopts and develops the 3WD theory to minimize the effects of the fuzzy feature selection process and obtain better performance. The features with uncertain relationships are put into the boundary region, determining the boundary of feature correlation division. The feature space is divided into three subspaces: strong relevance, weak relevance, and irrelevance. Then, different online redundancy analyzes are applied according to the correlation of features. Hence, incorporating the developed 3WD theory into fuzzy OS²FS is a highly interesting issue.

To this end, this article combines the principle of deep forest and 3WD theory and proposes a recursion-and-fuzziness reinforced online sparse streaming feature selection (RF-OS²FS). The RF-OS²FS model constructs the relationship between LFA-based estimates and the consequent feature selection process and deals with fuzzy feature selection, thereby implementing a highly efficient RF-OS²FS model. This model is mainly divided into two primary steps.

- The consequent feature selection process is embedded with LFA-based estimates to enhance the performance of missing entries completion recursively.
- extending and improving the 3WD theory and introducing it into OS²FS to reduce the negative impact of fuzzy feature selection.

By doing so, this article makes the following contributions.

- It proposes a multilevel structure to combine the consequent feature selection process with LFA-based estimates and enhance the performance of OS²FS.
- The streaming feature evaluation criteria of relevance and redundancy are improved based on the developed 3WD theory, which could improve the accuracy of OS²FS.

- Extensive experiments are conducted on ten real-world benchmark datasets to verify the proposed RF-OS²FS. The extensive experiments demonstrate the following:
 - the RF-OS²FS performs better than six existing OSFS and OS²FS methods in data sparsity scenarios;
 - the RF-OS²FS is effective in connecting two steps of OS²FS recursively and solving fuzzy feature selection problems;
 - the feasibility, validity, and stability of RF-OS²FS handling missing data is proved.

II. RELATED WORK

A. Online Streaming Feature Selection

The OSFS evaluates feature relevance and redundancy and selects the best subset from streaming features. Perkins and Theiler [19] first investigate the problem of the feature selection for streaming data and propose the Grafting algorithm by using a gradual gradient descent approach. However, Grafting constantly adjusts the regularization parameters of feature selection by using the known feature space. Then, in the work of Zhou et al. [20], Alpha-investing utilizes streamwise regression without requiring a global model. Wu et al. [21] constructed the OSFS framework, including OSFS and fast-OSFS, both of which have two steps: online relevance analysis and online redundancy analysis, but the fast-OSFS runs faster. Zhou et al. [22] proposed the OSSFS-DD model from the perspective of 3WD. Wang et al. [23] further developed the OSFS model by networked systems. Considering the relationship between features, Yu et al. [24] introduced the optimal streaming feature selection problem with the pairwise relationship between features under mutual information. The SFS-FI [25] method is present to consider the relationship between all the features. As for different types of streaming features, an OSFS framework for unknown type feature streams that adopted this exact scheme is constructed in [26].

There are also studies that focus on a rough set of handling streaming features [27]. Eskandari et al. [28] designed a novel step online significance analysis based on the rough set, namely OS-NRRSARA-SA. The K-OFSD model [29] adopts K nearest neighbors. Subsequently, Zhou et al. [30] used an adapted neighborhood rough set to evaluate the relevance, redundancy, and significance, named OFS-A3M. Zhou et al. [31] introduced an OFS-Density model based on adaptive density neighborhood relation. Li et al. [32] proposed the OFS-Gapknn for the feature selection problem of the gap neighborhood. The OSFS-ET model [33] is terminated early when it achieves the best effect. Zhang et al. [34] constructed a new measurement method for the feature flow in groups. Luo et al. [35] adopted a parallel process to propose the RHDOFS model. Despite existing OSFS models being essential in selecting streaming features on the fly, to the best of authors' knowledge, existing OSFS methods cannot efficiently deal with the sparse streaming features. Missing data can lead to an increase in the computational complexity of the OSFS model, while potentially selecting irrelevant and redundant features. Sparse streaming features may have weak correlations with other features or the target variable, making it difficult to accurately assess their relevance. In addition,

sparse streaming features can result in imbalanced data distributions, where certain sample values occur extremely infrequently, which can negatively impact the effectiveness of OSFS.

Note that the LFA model is a useful and meaningful method for pre-estimating the missing entries of a high-dimensional sparse matrix, which maps the sparse matrix into two latent factor matrices [36], [37]. Traditional methods, such as mean imputation and matrix factorization, fill missing values based on observed data patterns. These approaches often rely on assumptions like linearity or local similarity, making them less effective for capturing complex, nonlinear relationships in high-dimensional or sparse streaming data. In contrast, the LFA model captures the underlying data structure, enabling more effective handling of complex dependencies and nonlinear patterns through latent space modeling. The latest LFA model analyzes latent information based on graph neural networks [38] and matrix factorization [39], [40]. Recently, it has been successfully developed in an OS²FS scenario, i.e., the LOSSA algorithm [12]. It employs the LFA model to produce an approximation of the original sparse matrix. However, there is no connection between LFA-based estimates and the consequent feature selection process during OS²FS. The inevitable error exists between the completed value and the actual value, which causes an uncertain relationship between labels and completed sparse streaming features, and it leads to fuzziness.

B. Multilayered Structure Learning

Deep neural networks (DNNs) are powerful enough to fit complex tasks using multilevel structures. Each layer of DNNs consists of multiple neurons; each neuron receives the output of the previous level, and the processing result is transmitted to the next layer [41]. Due to a large amount of parameter adaptation, extensive training sets, and complicated calculation processes, DNNs consume more computing resources and time [42]. The deep forest makes use of DNNs to train multigrained cascade forests. It integrates the backbone of the decision tree and requires fewer parameters. Multigrained scanning and cascade forests are important parts of deep forests. Each level of the cascaded forest receives the feature information of the previous level and outputs the processing result to the next level. The cascade layer is determined by self-adjustment, so a few training samples can also obtain good performance. And each layer contains different types of forests to enhance diversity [43], [44].

Deep learning is a typical branch of the LFA model to construct a recursively-reinforced structure. For example, inspired by the deep forest, the PMLF model [14] overcomes the limitation of the LFA-based model and increases the density of known training data by generating synthetic data layer by layer. The MLF model [45] is widely adopted in deep forests to build deeply-structured models and extend LFA variants. The input information of each layer is generated randomly, and the parameter solution is continuously enhanced, which can save time during the training process. All in all, deep forest has a significant effect on the LFA model in sparse data problems. Hence, a recursively-reinforced multilevel LFA structure is constructed, which links the two steps of OS²FS, recursively

selecting features and generating synthetic data by complement sampling strategy to improve the representation ability.

3WD is one of the key technologies to handle various uncertain and fuzzy problems in the decision-making process, proposed by YAO [46], which puts uncertain objects into the boundary region to reduce the decision-making risk [47], [48]. Currently, it is widely used in the field of feature selection. Gao et al. [49] defined the measure of the parameterized maximum entropy and develop 3WD-based feature selection. Liu et al. [50] measured the neighborhood relevancy, redundancy, and granularity interactivity induced by neighborhood granularity interactivity. Qi et al. [51] established the three-way utility decision model for feature fuzziness in multiattribute environments. Fu et al. [52] used the measure of the 3WD-making to extend the multiattribute method with hesitant fuzzy information. Thus, feature selection based on the 3WD theory greatly improves efficiency. From this point of view, a 3WD theory becomes a potential solution to this uncertain relationship and fuzzy feature selection problem during OS²FS. In addition, some studies consider the fuzzy relationships between features. For example, Yin et al. [53] proposed a multiscale fusion-based fuzzy uncertainty measure to handle multilabel feature selection with missing labels. Furthermore, Yin et al. [54] introduced a robust multilabel feature selection method that takes into account fuzzy dependencies and feature interactions.

III. PRELIMINARIES

A. Online Streaming Feature Selection

The Nomenclature lists the adopted symbols of this article. The OSFS model offers a novel solution to selecting the optimal subset of streaming features, which addresses this problem by online relevance analysis and online redundancy analysis. Given a streaming feature set $F = \{F_1, F_2, \dots, F_T\}$, and the label set $C = [c_1, c_2, \dots, c_M]^T$ where $F_t = [f_{1,t}, f_{2,t}, \dots, f_{M,t}]^T$ have M simples, $t \in \{1, 2, \dots, T\}$.

Definition 1 (Conditional Independence [21]): Suppose that two features F_p and F_q , where $p \neq q$, $p, q \in \{1, 2, \dots, T\}$, if $P(F_p|F_q, X) = P(F_p, X)$ or $P(F_p|F_q, X) = P(F_q, X)$, such that F_p and F_q are conditionally independent to the subset $X \subseteq F$.

Definition 2 (Online Relevance Analysis [21]): For a streaming feature F_t at the time stamp t :

- a) if $\forall \varsigma \subseteq F - \{F_t\}$ s.t. $P(C|\varsigma, F_t) \neq P(C|\varsigma)$, then decide F_t is strong relevance;
- b) if $\exists \varsigma \subseteq F - \{F_t\}$ s.t. $P(C|\varsigma, F_t) \neq P(C|\varsigma)$, then decide F_t is weak relevance;
- c) if $\forall \varsigma \subseteq F - \{F_t\}$ s.t. $P(C|\varsigma, F_t) = P(C|\varsigma)$, then decide F_t is irrelevance.

Definition 3 (Online Redundancy Analysis [21]): Given a relevant feature F_t ($M(F_t) \notin M(C)_t$), and the redundant set $\{X_F\}$ is denoted as follows:

$$\begin{aligned} \forall X_F \in M(C)_t \cup F_t, \exists \varsigma \subseteq M(C)_t \cup F_t - \{X_F\} \\ \text{s.t. } P(C|X_F, \varsigma) = P(C|\varsigma) \end{aligned} \quad (1)$$

where $M(\cdot)$ denotes Markov blanket.

TABLE I
COST MATRIX

Action	Cost Function	
	X	X_C
a_P	λ_{PP}	λ_{PN}
a_B	λ_{BP}	λ_{BN}
a_N	λ_{NP}	λ_{NN}

B. Latent Factor Analysis

The LFA model is of great significance in pre-estimating the sparse matrix. This section first gives the definition of the LFA model as follows.

Definition 4 (LFA [36], [37]): Let $R^{M \times H}$ be a sparse matrix, and an LFA model trains two latent factor matrices $U^{M \times L}$ and $V^{H \times L}$ via the known entries, which precisely represent the rank- L approximation \hat{R} of R , where \hat{R} is formulated as $\hat{R} = UV^T$, L is the dimension of U and V , and $L \ll \min\{|M|, |H|\}$.

Then, the error formula of the LFA model is calculated as follows:

$$E(U, V) = \sum_{r_{m,h} \in \Lambda} e(\Delta_{m,h}), \Delta_{m,h} = r_{m,h} - \hat{r}_{m,h}(u_m, v_h) \quad (2)$$

where Λ denotes the known data of R , $e(\cdot)$ calculates the error between $r_{m,h}$ and $\hat{r}_{m,h}$, $r_{m,h}$ is m th row and h th column of R , the $\hat{r}_{m,h}$ is the predicted value for $r_{m,h}$, $\hat{r}_{m,h}(\cdot)$ stands for predictive function.

Regularization plays a crucial role in the LFA model to avoid overfitting. Therefore, the regularization is combined into (2), and the following objective function is obtained:

$$\begin{aligned} \varepsilon(U, V) = & \sum_{r_{m,h} \in \Lambda} e(r_{m,h} - \hat{r}_{m,h}(u_m, v_h)) \\ & + \frac{\lambda}{2} (\|U\|_F^2 + \|V\|_F^2) \end{aligned} \quad (3)$$

where $\|\cdot\|_F$ computes the Frobenius norm and λ represents the regularization coefficient.

C. Three-Way Decision

The 3WD theory focuses on the thought of thinking in “three” and divides the whole universe into three regions, s.t., positive, boundary, and negative, denoted as POS, BND, and NEG. YAO et al. [46] combined probabilistic rough sets with Bayesian decision theory to calculate the optimal α and β . That is, given two states set and three actions set: $\omega = \{X, X_C\}$ and $A = \{a_P, a_B, a_N\}$, where X and X_C represent belonging and not belonging, respectively, and a_P, a_B , and a_N represent accepted, delayed, and rejected decision, respectively. Different actions incur different costs, as shown in Table I. For example, λ_{PP} and λ_{PN} represent the costs of taking action a_P when x belongs to X and when it belongs to X_C , respectively. The decision rules for the 3WDs are as follows:

- a) if $P(X|[x]) \geq \alpha$, decide $x \in \text{POS}(X)$;
- b) if $\beta < P(X|[x]) < \alpha$, decide $x \in \text{BND}(X)$;
- c) if $P(X|[x]) \leq \beta$, decide $x \in \text{NEG}(X)$;

where $0 \leq \beta < \alpha \leq 1$, α and β are formulated as

$$\begin{aligned} \alpha &= \frac{(r_{PN} - r_{BN})}{(r_{PN} - r_{BN}) + (r_{BP} - r_{PP})} \\ \beta &= \frac{(r_{BN} - r_{NN})}{(r_{BN} - r_{NN}) + (r_{NP} - r_{BP})}. \end{aligned} \quad (4)$$

IV. PROPOSED ALGORITHM

A. Problem of RF-OS²FS

The set of sparse streaming features $F' = \{F'_1, F'_2, \dots, F'_T\}$ is assumed to be with the missing data rate ρ , where $\rho = 1 - |\Lambda|/M$, $|\cdot|$ stands for cardinal of the set. Sparse streaming features $F'_t, F'_{t+1}, \dots, F'_{t+H-1}$ are being generated one by one and flowing in $B^{M \times H}$ buffer from time point t to $t+H-1$, which is named as the sparse streaming feature matrix $B_t = \{F'_t, F'_{t+1}, \dots, F'_{t+H-1}\}$, where the size of the buffer is H . Then, the completed streaming feature matrix is estimated based on observed known data expressed as $\hat{B}_t = \{\hat{F}'_t, \hat{F}'_{t+1}, \dots, \hat{F}'_{t+H-1}\}$.

The primary aim of RF-OS²FS is to select the optimal features. Utilizing the concepts of deep forest can connect LFA-based estimates and consequent feature selection process and recursively enhance data density. At the same time, following the core idea of 3WD, the RF-OS²FS reduces the impact of the uncertainty between labels and completed sparse streaming features and fuzziness. Therefore, the RF-OS²FS needs to overcome the following issues:

$$S_t = \arg \max_{\sigma \subseteq S_t} P(C|\sigma). \quad (5)$$

B. Framework of RF-OS²FS

Following the core ideas of the deep forest [41] and the 3WD [46], the RF-OS²FS model is designed to recursively connect LFA-based estimates and consequent feature selection process and handle fuzzy feature selection. Our proposed method introduces a multilevel completion mechanism that iteratively estimates missing values. At each level, the model utilizes selected feature information to progressively enhance the accuracy of the completed values. This multilevel approach enables our method to better handle streaming data and capture underlying features that single-step LFA models might miss. In addition, the multilevel framework integrates the 3WD theory to effectively address fuzziness. At time stamp t , its training process for one layer is shown in Fig. 1(a), sequentially connecting multiple layers of training processes to construct a multilayer structure. The multilevel feature selection process is as follows:

- a) inputting $F'_t, F'_{t+1}, \dots, F'_{t+H-1}$ into the buffer;
- b) training the known data set Λ_t and the enhanced entries set Ω_t to estimate $\hat{B}_{t,d}$ only in parallel by nothing that $\Omega_t = \emptyset$ when $d = 1$;
- c) selecting optimal current selected features $S_{t,d}$ from $\hat{B}_{t,d}$;
- d) inputting the features (not in the S_t set but in the $S_{t,d}$ set) that enhance the correlation into S_t and Ω_t [i.e., the reinforced completed features are not found in Λ_t , then it is filled into Ω_t , as shown by the yellow module in Fig. 1(b)];

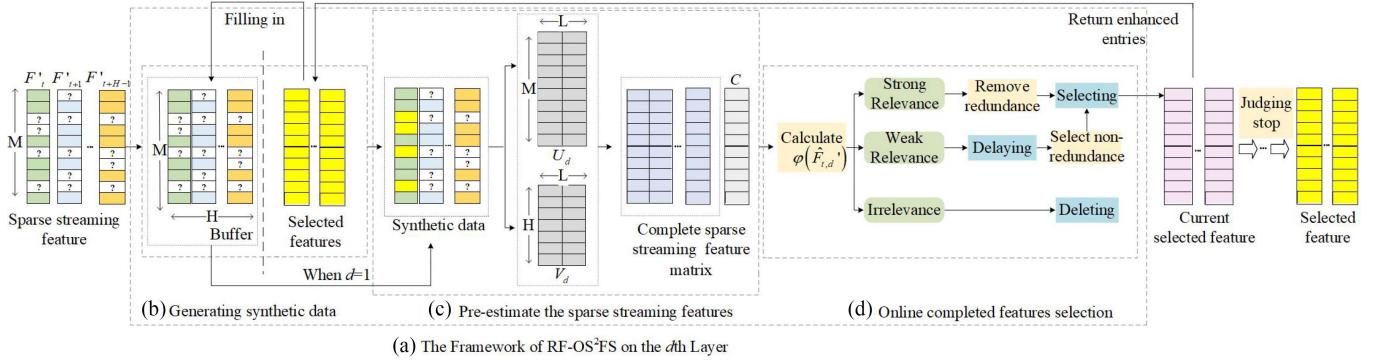


Fig. 1. Flowchart of RF-OS²FS model. (a) Framework of RF-OS²FS on the d th layer. (b) Generating synthetic data. (c) Pre-estimate the sparse streaming features. (d) Online completed features selection.

- e) repeating 2–4 until there are no new selected features or features cannot enhance the correlation;
- f) repeating 1–5 until no streaming feature flows in;
- g) output the optimal features subset S_t .

The recursively-reinforced model cannot go on indefinitely and needs to stop at the most suitable termination layer. This multilevel stop condition is to define an appropriate function to calculate the correlation enhancement for each layer.

Definition 5 (Correlation Enhancement Function): Given a feature F'_t and S_{t-1} , comparing the correlation of selected features in different layers, then the correlation enhancement function can be defined as

$$\text{DEP}(F'_t) < \psi \frac{1}{|S_{t-1}|} \sum_{F_\Phi \in S_{t-1}} \text{DEP}(F_\Phi) \quad (6)$$

where $\text{DEP}(\cdot)$ computes the correlation of F'_t and ψ is a function's parameter, respectively. At each layer, the correlation enhancement function quantifies the improvement in feature relevance when F'_t is added to S_{t-1} . The recursive process terminates when the correlation enhancement value $\text{DEP}(F'_t)$ falls below a predefined threshold, indicating that further layers do not significantly improve feature relevance.

C. Dth Layer of RF-OS²FS

The RF-OS²FS is a recursively-reinforced structure for connecting two steps of OS²FS dealing with fuzzy feature selection problem. On the d th layer, the synthetic data is first generated by the reinforced completed features. The missing values are estimated, and the 3WD is promoted and expanded to select sparse streaming features.

1) Generating Synthetic Data: Given a sparse streaming feature matrix $B_t = \{F'_t, F'_{t+1}, \dots, F'_{t+H-1}\}$, it is very crucial to increase its known data density to improve the representation learning ability of the LFA model. Following this principle, RF-OS²FS increases its data density by selecting reinforced completed features layer by layer. Let $\hat{B}_{t,d}$ be the input matrix of the d th layer, in which the $\hat{B}_{t,d}$ contains all known data elements of Λ_t , and the completion sampling strategy of generating synthetic data at the d th layer is as follows:

- a) inputting reinforced completed features to reset its synthetic data;
- b) finding elements in reinforced completed features that are not in Λ_t [as shown by the yellow module in Fig. 1(b)] and putting such elements in Ω_t ;
- c) outputting the synthetic data, which contains known data and enhanced entries;
- d) repeating Steps 1–3 until the recursively-reinforced multilevel process terminates.

2) Pre-estimate the Sparse Streaming Features: This article designs a recursively-reinforced structure for selecting reinforced completed features on each layer to increase data density, which can enhance the representation ability. The process of recursively-reinforced multilevel feature completion is explained in detail, and the d ($d \in \{1, 2, \dots, D\}$) layer at timestamp t is taken as an example, as shown in Fig. 1(c). Both matrices U and V are initialized with small random values, where each element is derived from a random permutation scaled to a range close to zero. The initialization process ensures that the matrices start with nonzero but small initial values, which can be useful in iterative algorithms where initial conditions matter. The update method is illustrated below using U as an example

$$u_{m,k} = 0.004 - 0.004 \times \frac{\text{randperm}(1000, 1)}{1000}. \quad (7)$$

Note that when $d = 1$, the missing value is completed only based on the known dataset Λ_t . However, when $d > 1$, on this basis, the enhanced entries set Ω_t can be used. Extending from (3), the objective function is optimized by stochastic gradient descent as follows:

$$\begin{aligned} \forall f'_{m,j} \in \Lambda_t \quad \text{or} \quad \Omega_t \\ \left\{ \begin{aligned} u_m^d &\leftarrow u_m^d - \eta \frac{\partial \varepsilon(U^d, V^d)}{\partial u_m^d} \\ v_j^d &\leftarrow v_j^d - \eta \frac{\partial \varepsilon(U^d, V^d)}{\partial v_j^d} \end{aligned} \right. \end{aligned} \quad (8)$$

where η stands for the learning rate, the completed sparse streaming feature matrix is formulated as $\hat{B}_{t,d} = U_d V_d^\top (\hat{B}_{t,d} = \{\hat{F}'_{t,d}, \hat{F}'_{t+1,d}, \dots, \hat{F}'_{t+H-1,d}\})$, and the training rules for two

latent factor matrices are constructed as follows:

$$\begin{aligned}
 d > 1 \quad \forall f'_{m,j} \notin \Lambda_t, \quad \text{and} \quad f'_{m,j} \in \Omega_t \\
 u^d_{m,k} &\leftarrow u^d_{m,k} + \theta v^d_{j,k} \eta \left(f'_{m,j} - \sum_{k=1}^L u^d_{m,k} v^d_{j,k} \right) - \lambda \eta u^d_{m,k} \\
 v^d_{j,k} &\leftarrow v^d_{j,k} + \theta u^d_{m,k} \eta \left(f'_{m,j} - \sum_{k=1}^L u^d_{m,k} v^d_{j,k} \right) - \lambda \eta v^d_{j,k} \\
 d > 1 \quad \forall f'_{m,j} \notin \Lambda_t, \quad \text{and} \quad f'_{m,j} \in \Omega_t \\
 u^d_{m,k} &\leftarrow u^d_{m,k} + \theta v^d_{j,k} \eta \left(f'_{m,j} - \sum_{k=1}^L u^d_{m,k} v^d_{j,k} \right) - \lambda \eta u^d_{m,k} \\
 v^d_{j,k} &\leftarrow v^d_{j,k} + \theta u^d_{m,k} \eta \left(f'_{m,j} - \sum_{k=1}^L u^d_{m,k} v^d_{j,k} \right) - \lambda \eta v^d_{j,k}. \quad (9)
 \end{aligned}$$

In which $k \in \{1, 2, \dots, L\}$, $m \in \{1, 2, \dots, M\}$, $j \in \{t, t+1, \dots, t+H-1\}$, θ stands for the parameter of enhanced data, the entry $u^d_{m,k}$ at the m th row and k th column of U_d , and the entry $v^d_{j,k}$ at the j th row and k th column of V_d .

3) *Online Completed Features Selection*: The RF-OS²FS structure completed sparse streaming features can improve completion efficiency. However, there are still fewer errors between the completed features and the actual values, which causes the uncertain relationships between the sparse features and the labels. For handling uncertain information, 3WD is more efficient than traditional two-way decisions because the key aspect of 3WD is the ability to make delayed decisions, thereby reducing the losses caused by incorrect decisions [55]. Because of the uncertainty of the sparse features and labels, the boundary of correlation division of each feature is unclear and gradually changes, which leads to the existence of fuzziness. The 3WD theory can be extended and promoted, and the loss function of each feature can be established, and each feature has a different boundary of correlation division.

Therefore, leveraging the concept of 3WD [46], the RF-OS²FS model adopts a delayed decision, categorizing features with uncertain relationships as weak relevance, thus reducing the influence of misclassification, as shown in Fig. 1(d). This article constructs the state space $\omega = \{X, X_C\}$ for sparse streaming features, where X represents relevant features, and X_C represents irrelevant features. The decision set $A = \{a_P, a_B, a_N\}$ divides the features into strongly relevant, weakly relevant, and irrelevant features, respectively. At time stamp t and level d , the evaluation function $\varphi(\hat{F}'_{t,d})$ measures the degree of fuzziness in feature relevance for completed streaming feature matrix $\hat{B}_{t,d}$ within a feature $\hat{F}'_{t,d}$:

- a) when $\varphi(\hat{F}'_{t,d}) \geq \alpha$, $\hat{F}'_{t,d}$ is classified into the positive region of the set $\hat{U}_{t,d}$, which represents strongly relevant features;
- b) when $\beta < \varphi(\hat{F}'_{t,d}) < \alpha$, $\hat{F}'_{t,d}$ is classified into the boundary region of the set $\hat{U}_{t,d}$, which represents weakly relevant features;

c) when $\varphi(\hat{F}'_{t,d}) \leq \beta$, $\hat{F}'_{t,d}$ is classified into the negative region of the set $\hat{U}_{t,d}$, which represents irrelevant features.

Based on this, the $\hat{U}_{t,d}$ is dynamically divided into three mutually exclusive feature sets, expressed as

$$\begin{aligned}
 \text{POS} &= \left\{ \hat{F}'_{t,d} \in \hat{U}_{t,d} \mid \varphi(\hat{F}'_{t,d}) \geq \alpha \right\} \\
 \text{BND} &= \left\{ \hat{F}'_{t,d} \in \hat{U}_{t,d} \mid \beta < \varphi(\hat{F}'_{t,d}) < \alpha \right\} \\
 \text{NEG} &= \left\{ \hat{F}'_{t,d} \in \hat{U}_{t,d} \mid \varphi(\hat{F}'_{t,d}) \leq \beta \right\}. \quad (10)
 \end{aligned}$$

Different decision-making methods result in different loss values, as shown in Table I. For example, when feature $\hat{F}'_{t,d}$ belongs to X , λ_{PP} , λ_{BP} , and λ_{NP} represent the corresponding costs when the feature is classified into the sets $\text{POS}_{(\alpha,\beta)}(X)$, $\text{BND}_{(\alpha,\beta)}(X)$, and $\text{NEG}_{(\alpha,\beta)}(X)$, respectively. The threshold parameters α and β are derived from six cost values. For these six cost values, correct classification incurs no cost, i.e., $\lambda_{NN} = \lambda_{PP} = 0$. Thus, the threshold parameters α and β are only related to the remaining four cost functions. The total risk cost associated with feature classification is given by

$$\begin{aligned}
 \text{cost}_{(\alpha,\beta)} &= \sum_{\hat{F}'_{t,d} \in \text{POS}_{(\alpha,\beta)}(X)} \lambda_{PN} + \sum_{\hat{F}'_{t,d} \in \text{BND}_{(\alpha,\beta)}(X)} \lambda_{BP} \\
 &+ \sum_{\hat{F}'_{t,d} \in \text{BND}_{(\alpha,\beta)}(X_C)} \lambda_{BN} + \sum_{\hat{F}'_{t,d} \in \text{NEG}_{(\alpha,\beta)}(X)} \lambda_{NP}. \quad (11)
 \end{aligned}$$

The optimization problem with the goal of minimizing loss can be expressed as

$$\begin{aligned}
 \min \text{cost}_{(\alpha,\beta)} &= \sum_{\varphi(\hat{F}'_{t,d}) \geq \alpha} \lambda_{PN} + \sum_{\beta < \varphi(\hat{F}'_{t,d}) < \alpha} \lambda_{BP} \\
 &+ \sum_{\beta < \varphi(\hat{F}'_{t,d}) < \alpha} \lambda_{BN} + \sum_{\varphi(\hat{F}'_{t,d}) \leq \beta} \lambda_{NP}. \quad (12)
 \end{aligned}$$

In practical application, risk parameters are often determined subjectively according to experts' experience and set according to experience. Therefore, finding the optimal thresholds becomes a crucial problem. According to (12), the RF-OS²FS model transforms the optimal threshold into an optimization problem. Solving this optimization problem yields the thresholds that minimize decision risk costs. Using the simulated annealing algorithm, let $f = \text{cost}_{(\alpha,\beta)}$ serve as the fitness function, with an initial temperature set to 100. The temperature changes according to $T_{i+1} = r \cdot T_i$, where i is the iteration number and r is a constant used for temperature regulation, set to 0.95 [56].

Different redundancy analysis methods are applied to different features. Strongly relevant features are directly placed into S_t and then checked for redundancy, i.e.,

$$\exists X_R \in S_t \quad \text{s.t.} \quad P(C \mid \hat{F}'_{t,d}, X_R) = P(C \mid X_R). \quad (13)$$

Weakly relevant features, on the other hand, must be checked for redundancy against the features in S_t (provided S_t is not empty) before deciding whether to place them into S_t , i.e.,

$$\forall X_R \in S_t \quad \text{s.t.} \quad P(C \mid \hat{F}'_{t,d}, X_R) \neq P(C \mid X_R). \quad (14)$$

TABLE II
CONDITIONAL PROBABILITY

Action	Conditional probability	
	X	X_C
a_P	Pc_{PP}^t	Pc_{PN}^t
a_B	Pc_{BP}^t	Pc_{BN}^t
a_N	Pc_{NP}^t	Pc_{NN}^t

D. Algorithm Design and Time Complexity Analysis

Algorithm design: In order to improve the ability to deal with sparse streaming features, this article proposes a recursively-reinforced fuzzy OS²FS model, as shown in Algorithm 1. In steps 3–6, features flow into the buffer, and then a recursively-reinforced training process is carried out in steps 7–35. At layer d , the missing values are estimated at the known values of Λ_t and Ω_t in steps 9–15. In steps 16–27, feature selection is carried out for the completed streaming feature matrix with minimum decision risk. It is judged whether to stop multilevel training in steps 28–34.

Time complexity analysis: Compared to [10], the multilevel time complexity is $O(D)$. At the d layer, the time complexity of completing sparse streaming features is $O(M \times T \times (1-\rho) \times L)$. Selecting completed sparse streaming features has a time complexity of $O(H)$ to evaluate the uncertainty.

E. Detailed Empirical Studies Regarding RF-OS²FS's Cost

Considering the probability that the feature belongs to POS, BND, and NEG, the conditional probability of the feature in six states is shown in Table II. Therefore, the total cost of the three-way feature classification can be formulated as

$$\begin{aligned} E\text{cost}_{(\alpha,\beta)} = & r_{NP}Pc_{NP}^t + r_{PN}Pc_{PN}^t \\ & + r_{BP}Pc_{BP}^t + r_{BN}Pc_{BN}^t. \end{aligned} \quad (15)$$

The traditional two-way decision only includes accepting the rejection decision, not the delayed decision [57], [58]. Similarly, the cost of correct two-way feature classification is zero, i.e., $\lambda_{NN} = \lambda_{PP} = 0$. Based on (15), the total expected cost of two-way feature classification is calculated by

$$E\text{cost}_{(\alpha,\beta)}^{(2)} = r_{PN}^{(2)}Pc_{PN}^{t,(2)} + r_{NP}^{(2)}Pc_{NP}^{t,(2)}. \quad (16)$$

And then, three-way classification divides uncertain features into boundary region, while certain features do not change [55], i.e.,

$$\begin{aligned} Pc_{PN}^{t,(2)} &= Pc_{PN}^t + Pc_{BN}^t \\ Pc_{NP}^{t,(2)} &= Pc_{NP}^t + Pc_{BP}^t \\ Pc_{PP}^{t,(2)} &= Pc_{PP}^t \\ Pc_{NN}^{t,(2)} &= Pc_{NN}^t. \end{aligned} \quad (17)$$

Proposition 1: The total expected cost of the two-way feature classification is more than that of the three-way, i.e., $\text{cost}_{(\alpha,\beta)} < \text{cost}_{(\alpha,\beta)}^{(2)}$.

Algorithm 1: The RF-OS²FS model.

```

Initialize  $H, \theta, \psi$ 
repeat
  a new sparse streaming feature  $F'_t$  arrived at timestamp  $t$ 
  if  $N \neq 0$ 
     $N = N - 1$ ,  $B_t = B_t \cup F'_t$ 
  endif
   $d = 1$ 
  repeat
    while  $\Gamma_T = \text{Max}$ 
      for  $j=t$  to  $t+H-1$ 
        for  $\forall f'_{m,j} \in O_t$  or  $K_t$ 
          Estimate  $u_{m,k}^d, v_{j,k}^d$  according to (8),  $\Gamma_T = \Gamma_T - 1$ 
        end for
      end for
    end while  $\hat{B}_{t,d} = U_d V_d^T$ 
    for  $i=1$  to  $P$ , fetch  $\hat{F}'_{t+i,d}$  from  $\hat{B}_{t,d}$ , calculate  $\alpha_{t+d}, \beta_{t+d}$ 
    if  $\varphi(\hat{F}'_{t+i,d}) \leq \beta_{t+d}$ , discard  $\hat{F}'_{t+i,d}$ 
    else if  $\varphi(\hat{F}'_{t+i,d}) \geq \alpha_{t+d}$ 
       $S_{t,d} = S_{t,d} \cup \hat{F}'_{t+i,d}$  following (13), and update  $\Omega_t$ 
    else
      if  $S_{t,d} \neq \emptyset$ , accept  $\hat{F}'_{t+i,d}$  following (14)
      else  $BND = BND \cup \hat{F}'_{t+i,d}$ 
      accept  $\hat{F}'_{t+i,d} \in BND$  following (14) until  $S_{t,d} \neq \emptyset$ 
    end if
  end if
  Remove redundant features in  $S_{t,d}$  according to Definition 1,  $i=i+1$ 
  end for
  if  $\exists \hat{F}'_{t+\sigma,d} \in S_{t,d}$  and  $\hat{F}'_{t+\sigma,d} \notin S_t$ 
    if  $\hat{F}'_{t+\sigma,d}$  satisfies (6)
      put  $\hat{F}'_{t+\sigma,d}$  into  $S_t$ 
    else terminate
    end if
  else terminate
  end if
   $d=d+1$ 
until no features are available
Output  $S_t$ 

```

Proof: Based on (17), we have

$$\begin{aligned} E\text{cost}_{(\alpha,\beta)} - E\text{cost}_{(\alpha,\beta)}^{(2)} &= (r_{NP}Pc_{NP}^t + r_{PN}Pc_{PN}^t + r_{BP}Pc_{BP}^t + r_{BN}Pc_{BN}^t) \\ &\quad - (r_{PN}Pc_{PN}^{t,(2)} + r_{NP}Pc_{NP}^{t,(2)}) \\ &= r_{NP} * (Pc_{NP}^t - Pc_{NP}^{t,(2)}) + r_{PN} * (Pc_{PN}^t - Pc_{PN}^{t,(2)}) \\ &\quad + r_{BP}Pc_{BP}^t + r_{BN}Pc_{BN}^t \\ &= Pc_{BP}^t * (r_{BP} - r_{NP}) + Pc_{BN}^t * (r_{BN} - r_{PN}). \end{aligned}$$

TABLE III
DETAILS OF SELECTED DATASETS

Mark	Dataset	Features	Instances	Class
D1	Colon	2001	62	2
D2	Lung	3313	83	5
D3	lymphoma	4027	96	9
D4	Brain-Tumor1	5921	90	5
D5	Prostate	6034	102	2
D6	Leukemia	7071	72	2
D7	ALLAML	7130	72	2
D8	Lungcancer	12534	181	2
D9	MLL	12583	72	3
D10	GLI_85	22284	85	2

TABLE IV
ALL THE ALGORITHM PARAMETERS USED IN THE EXPERIMENTS

Mark	Algorithm	Parameter
M1	RF-OS ² FS	Z test, Alpha is 0.05.
M2	LOSSA	Z test, Alpha is 0.05. (TSMC, 2022)
M3	Fast-OSFS	Z test, Alpha is 0.05. (TPAMI, 2013)
M4	SAOLA	Z test, Alpha is 0.05. (TKDD, 2016)
M5	SFS-FI	Z test, Alpha is 0.01, $\gamma=0.01$. (TNNLS, 2021)
M6	OFS-A3M	(Information Sciences, 2019)
M7	RHOFS	(TPDS, 2023)

TABLE V
ALL THE CLASSIFIER PARAMETERS USED IN THE EXPERIMENTS

Classifier	Parameter
KNN	The neighbors are 3.
RF	6 decision trees.
CART	Default parameters values.

Since $r_{BP} < r_{NP}$ and $r_{BN} < r_{PN}$, then $\text{cost}_{(\alpha, \beta)} < \text{cost}_{(\alpha, \beta)}^{(2)}$. Compared with two-way features classification, the cost of three-way classification is less. Hence, Proposition 1 holds.

V. EXPERIMENTS AND RESULTS

In the section, the experiments aim to answer these issues as follows.

- 1) R.Q. 1. Does RF-OS²FS outperform state-of-the-art OSFS and OS²FS models in missing data scenario?
- 2) R.Q. 2. How does the increasing missing data rate impact the performance of all these models?
- 3) R.Q. 3. Do the parameters influence RF-OS²FS's average accuracy?

A. General Settings

Baselines: Experiments were conducted on ten real-world benchmark datasets to verify the performance of RF-OS²FS, LOSSA, and five OSFS models (Fast-OSFS, SAOLA, SFS-FI, OFS-A3M, and RHOFS). The details of these data sets and models are described in Tables III and IV, respectively. All algorithms are conducted on MATLAB. The experiments used three types of classifiers to measure the efficiency, i.e., support vector machine, random forest (RF), and classification and regression tree (CART) (see Table V).

Experimental designs: These experiments employ five-fold cross-validation, i.e., each dataset is randomly divided into two parts: training set (80%), and test set (20%). Training these

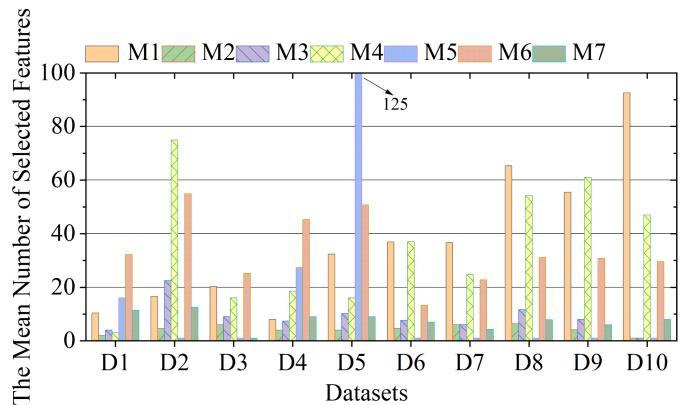


Fig. 2. Mean number of selected features varying with different algorithms.

datasets ten times and then recording the results of accuracy. To evaluate the obtained results, the Friedman test that adopts the 95% significance level is performed under the null hypothesis, which confirms whether the RF-OS²FS is significantly different from others. The Wilcoxon signed-ranks test is conducted to analyze whether the RF-OS²FS is more advantageous than all these competitive models [59]. All the experiments are carried out on Windows 10 with an Intel i7 2.40GHz CPU and 16GB RAM.

B. RF-OS²FS Versus OS²FS and OSFS Models (R.q. 1)

This experiment compares RF-OS²FS with LOSSA and five state-of-the-art traditional OSFS models on ten data sets when the missing rate is 0.1. The missing data of OSFS models can be revised by filling in zero. The experimental results of OS²FS and OSFS models better than others in the table are highlighted.

1) *Selected Features Analysis:* Fig. 2 shows the number of selected features in seven different algorithms on the ten data sets. From the Fig. 2, we observe as follows.

- a) For traditional OSFS models, the OSFS model selects fewer features than the RF-OS²FS on some datasets. Especially, SFS-FI only selects one feature on D2, 3, 6–10 that may cause the loss of important information. On D4, the SFS-FI algorithm obtains the most features. It shows that it is unstable when dealing with sparse streaming features. However, for some datasets, the OSFS model selects more features with lower average accuracy in the following experimental results, denoting that the selected features of traditional OSFS models contain irrelevance and redundancy in incomplete dataset scenarios.
- b) The RF-OS²FS selects slightly more features than LOSSA on ten datasets. With an increase in the number of features, the RF-OS²FS algorithm can improve accuracy that is compared to LOSSA from the following accuracy result. The RF-OS²FS model employs an iterative process to complete and select features multiple times, ensuring that all potential features are thoroughly evaluated and selected. This iterative approach allows RF-OS²FS to capture a more comprehensive set of relevant features, even if the number of selected features is slightly higher. In

TABLE VI
DETAILS OF SELECTED DATASETS USING THE SELECTED FEATURES TO TRAIN THREE CLASSIFIERS AND THEN TESTING ITS ACCURACY (%) WHEN THE MISSING DATA RATE IS 0.1

M D	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	Average	¹ Rank
M1	82.07 \pm 3.02	87.66 \pm 1.66	76.14 \pm 2.84	75.93 \pm 5.24	87.32 \pm 2.83	95.19 \pm 2.22	94.19 \pm 2.33	97.68 \pm 0.85	83.98 \pm 0.51	83.53 \pm 2.77	86.37 \pm 2.43	1.00
M2	80.45 \pm 2.59	84.79 \pm 2.77	67.56 \pm 3.28	72.30 \pm 2.40	86.64 \pm 1.61	94.95 \pm 2.22	92.87 \pm 1.71	97.41 \pm 0.70	81.01 \pm 3.03	67.02 \pm 2.76	82.50 \pm 2.31	3.00
M3	78.88 \pm 2.59	84.40 \pm 2.43	56.67 \pm 2.79	71.41 \pm 3.88	87.05 \pm 2.46	94.97 \pm 2.26	92.47 \pm 2.23	97.00 \pm 0.89	79.29 \pm 3.24	63.29 \pm 3.53	80.54 \pm 2.63	3.90
M4	77.65 \pm 3.19	83.38 \pm 2.32	61.05 \pm 5.16	71.37 \pm 3.17	86.23 \pm 2.80	93.53 \pm 1.61	93.14 \pm 2.18	96.10 \pm 0.84	80.15 \pm 3.83	82.75 \pm 1.66	82.54 \pm 2.68	4.30
M5	78.63 \pm 2.66	62.49 \pm 2.96	45.57 \pm 1.06	60.59 \pm 5.96	83.26 \pm 7.68	65.95 \pm 2.97	62.45 \pm 5.53	83.21 \pm 2.00	47.53 \pm 3.83	62.22 \pm 4.98	65.19 \pm 3.96	6.70
M6	74.07 \pm 5.52	85.47 \pm 3.11	71.03 \pm 5.22	68.30 \pm 4.10	79.81 \pm 5.46	88.53 \pm 4.23	83.00 \pm 5.81	92.99 \pm 2.29	75.89 \pm 6.80	73.86 \pm 7.60	79.30 \pm 5.01	5.30
M7	79.82 \pm 2.68	86.78 \pm 3.00	49.16 \pm 1.78	71.22 \pm 3.80	85.32 \pm 3.61	94.08 \pm 1.68	93.47 \pm 2.96	96.13 \pm 1.18	77.37 \pm 4.52	83.13 \pm 0.55	81.65 \pm 2.58	3.80

¹ Denoting the Average rank. The bold values stand for high performance.

contrast, LOSSA performs feature completion only once, which leads to larger completion errors and less accurate feature selection. As a result, LOSSA tends to select fewer features, but this comes at the cost of potentially missing important features that could improve model performance.

To verify the availability of the selected features, Table VI presents the average accuracy obtained by KNN, RF, and CART on all datasets.

The Friedman test: The ρ -values of the Friedman test in Table VI is $1.2748e - 08$. Hence, there is a significant difference between RF-OS²FS and other models.

The average accuracy of selected features: From Table VI, we can find the results as follows.

- a) RF-OS²FS is significantly greater than other models. It achieves a lower average rank than others, and the LOSSA is second only to it on most datasets, indicating that the effect of the LFA model to complete missing data is better than that of the traditional algorithm. Therefore, the traditional OSFS model is not reliable on incomplete datasets. It is mainly because of the low completion accuracy of missing values that important feature information is missing. The LFA-based model improves the efficiency of missing entries.
- b) LOSSA shows a higher precision than five OSFS models but lower than RF-OS²FS on most datasets. This proves that the single LFA model can improve the accuracy of feature selection. However, on D10, the accuracy of LOSSA is not as high as that of M4 and M7, and its performance is a little unstable. On this basis, RF-OS²FS selects optimal features and overcomes fuzzy feature selection problem to improve the effect of OS²FS further.

Wilcoxon signed-ranks test: The ρ -values of the Wilcoxon signed-ranks test between RF-OS²FS and others are provided in Table VII, both of which are less than 0.1, indicating that the RF-OS²FS is more efficient than that of others.

C. Influence of Missing Data Rate (R.q. 2)

In terms of the impact of missing data rate, the experiments were conducted on ten datasets with missing rates set at 0.3, 0.5, 0.7, and 0.9. Table VIII lists the ρ -values of the Wilcoxon signed-ranks test results. Fig. 3 shows the average accuracy of the seven algorithms on each dataset. From the center of the circle to the outermost circle, the accuracy ranges from 0 to 100, and the accuracy of each layer increases by 10. From this, we can observe the following.

- a) The accuracy of feature selection of all algorithms decreases with the increase of data missing rate on each

TABLE VII
STATISTICAL RESULTS OF THE WILCOXON SIGNED-RANK TEST IN TABLES VII
(ON ALL DATASETS)

RF-OS ² FS versus other algorithms	² R+	² R-	³ ρ -values
M2	55	0	9.7656e-04
M3	55	0	9.7656e-04
M4	55	0	9.7656e-04
M5	55	0	9.7656e-04
M6	55	0	9.7656e-04
M7	55	0	9.7656e-04

² A larger value denotes a higher accuracy.

³ There is no significant difference when ρ -values $\in [0.1, 0.9]$ at the 0.1 significance level.

dataset. There is a big drop in some data, but the missing rate has little effect on the accuracy in some data. On most data sets, the RF-OS²FS achieves the highest accuracy when the missing data rate is between 0.3 and 0.9. Two key components (recursively-reinforced selection structure and fuzzy feature selection) ensure the effectiveness of the RF-OS²FS framework.

- b) The accuracy of the LOSSA model is second only to RF-OS²FS when the missing data rate is between 0.1 and 0.7. The reason for this difference may be that few known values for estimating complete data with the increase of the missing data rate. RF-OS²FS is a recursively-reinforced method based on LOSSA that can significantly improve the representation ability of the LFA model. And RF-OS²FS effectively deals with fuzzy feature selection problem.

D. Analysis on Parameters (R.q. 3)

This section aims to analyze the influence of RF-OS²FS's parameters on the accuracy of feature selection. To this end, the experiments of parameters ψ , θ , and H were conducted.

- 1) *Analysis on Parameter ψ :* This experiment analyzes the influence of parameter ψ on the feature selection of the RF-OS²FS model, with ψ values set to 1.00, 1.25, 1.50, 1.75, and 2.00. Table IX records the average accuracy for different ψ values. For Table IX, the ρ -values of the Friedman test is $5.0871e - 07$, indicating that different ψ values have an impact on the RF-OS²FS model. Next, the most suitable ψ value will be selected for the RF-OS²FS model. From the experimental results in Table IX, it can be seen that $\psi = 1.25$ obtains the highest average and the lowest average rank. The OS²FS for the Wilcoxon signed-ranks tests for $\psi = 1.25$ compared to $\psi = 1.00$, 1.50, 1.75, and 2.00 are 0.0029, 0.0244, 0.0029, and 0.0068, respectively, demonstrating that $\psi = 1.25$ is better than the other ψ values. Therefore, it is recommended to set ψ at 1.25.

TABLE VIII
RANK SUM OF THE WILCOXON SIGNED-RANKS TEST ON OSFS AND OS²FS MODELS

ρ	M2		M3		M4		M5		M6		M7	
	$^2R+$	$^2R-$										
0.3	55	0	55	0	54	1	55	0	55	0	52	3
0.5	55	0	55	0	55	0	55	0	55	0	55	0
0.7	55	0	55	0	51	4	55	0	55	0	54	1
0.9	55	0	55	0	41	14	55	0	55	0	40	15

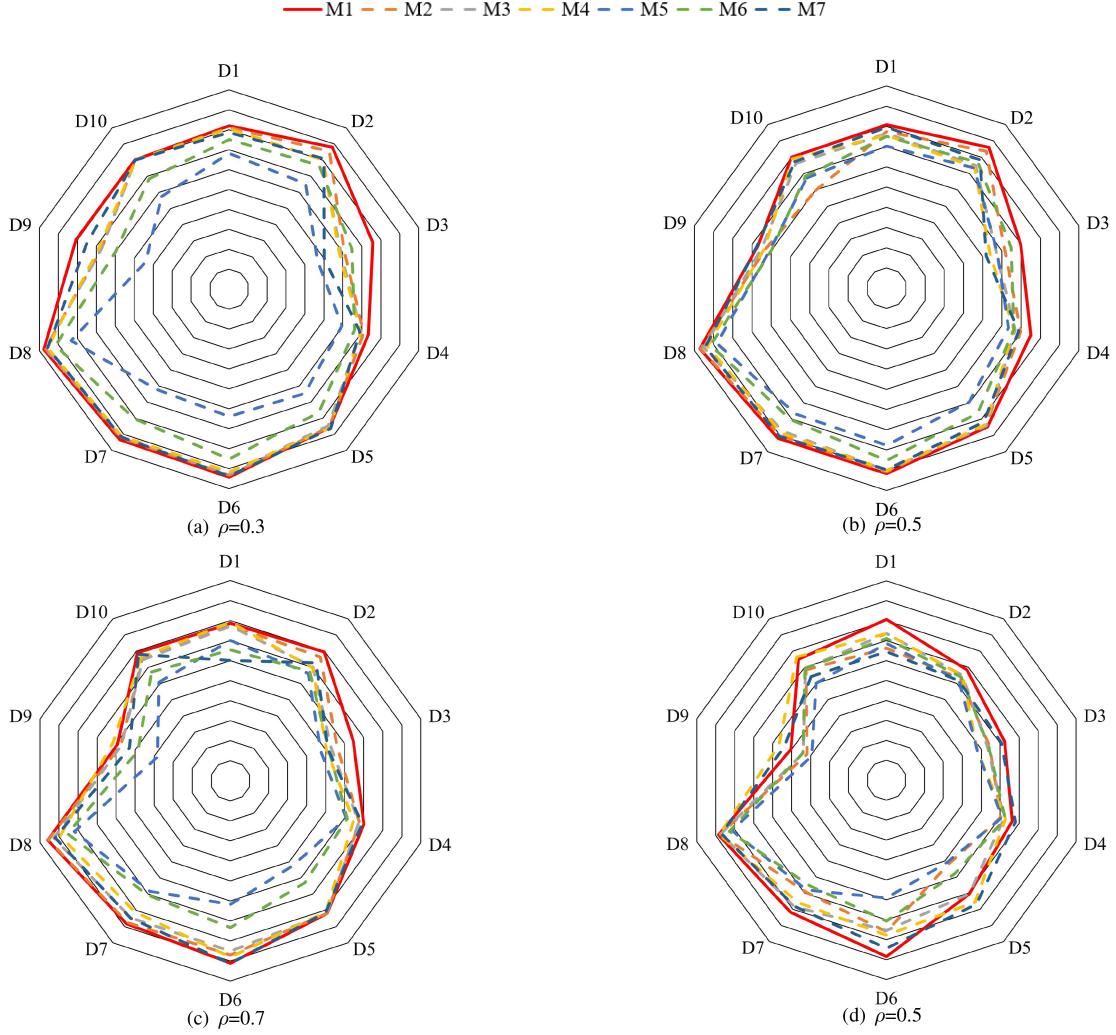


Fig. 3. Average accuracy comparison of OSFS and OS²FS models on each dataset.

TABLE IX
AVERAGE ACCURACY OF SELECTED FEATURES VARYING WITH DIFFERENT PARAMETERS OF THE MAPPING FUNCTION

M\D	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	Average	¹ Rank
1.00	82.74 \pm 2.18	86.87 \pm 2.35	72.74 \pm 1.29	70.74 \pm 1.57	86.79 \pm 1.61	93.94 \pm 1.30	93.21 \pm 1.03	96.79 \pm 0.38	73.44 \pm 2.27	82.63 \pm 1.53	83.99 \pm 1.55	3.70
1.25	82.07 \pm 3.02	87.66 \pm 1.66	76.14 \pm 2.84	75.93 \pm 5.24	87.32 \pm 2.83	95.19 \pm 2.22	94.19 \pm 2.33	97.68 \pm 0.85	83.98 \pm 0.51	83.53 \pm 2.77	86.37 \pm 2.43	1.60
1.50	81.47 \pm 1.84	87.93 \pm 1.75	67.62 \pm 2.02	71.67 \pm 4.98	86.78 \pm 1.19	93.51 \pm 1.95	95.65 \pm 1.71	97.79 \pm 0.26	75.29 \pm 4.99	80.20 \pm 4.71	83.80 \pm 2.54	3.20
1.75	79.21 \pm 5.83	84.98 \pm 0.84	71.68 \pm 3.62	74.63 \pm 0.79	85.95 \pm 3.19	94.43 \pm 0.70	94.71 \pm 2.31	97.34 \pm 0.67	82.63 \pm 1.53	80.20 \pm 4.71	84.58 \pm 2.42	3.60
2.00	79.57 \pm 3.08	85.86 \pm 3.71	73.75 \pm 1.38	74.07 \pm 4.71	86.94 \pm 5.04	93.52 \pm 2.69	94.00 \pm 1.93	97.61 \pm 0.52	84.78 \pm 3.34	81.18 \pm 2.22	85.13 \pm 2.86	3.00

The bold values stand for high performance.

TABLE X
AVERAGE ACCURACY OF SELECTED FEATURES VARYING WITH DIFFERENT CONTROLLING PARAMETERS

M\D	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	Average	¹ Rank
0.10	82.07 ± 2.87	85.46 ± 0.77	69.78 ± 2.92	77.78 ± 3.14	85.47 ± 1.18	95.65 ± 2.87	94.48 ± 3.28	97.61 ± 0.78	83.22 ± 2.94	79.02 ± 5.27	85.05 ± 2.60	3.10
0.25	83.14 ± 1.30	86.68 ± 1.65	70.12 ± 1.41	73.89 ± 0.79	87.95 ± 2.81	92.83 ± 2.96	92.41 ± 2.24	97.32 ± 0.13	83.87 ± 4.04	77.45 ± 5.27	84.57 ± 2.26	3.20
0.50	82.07 ± 3.02	87.66 ± 1.66	76.14 ± 2.84	75.93 ± 5.24	87.32 ± 2.83	95.19 ± 2.22	94.19 ± 2.33	97.68 ± 0.85	83.98 ± 0.51	83.53 ± 2.77	86.37 ± 2.43	1.80
0.75	79.83 ± 3.51	86.46 ± 0.81	74.10 ± 2.17	72.96 ± 3.67	86.76 ± 5.30	94.95 ± 0.67	92.33 ± 1.01	97.52 ± 0.39	80.83 ± 8.75	76.67 ± 3.05	84.24 ± 2.93	4.10
1.00	80.71 ± 1.60	87.36 ± 2.06	72.18 ± 4.40	73.15 ± 3.40	86.62 ± 0.99	92.30 ± 2.40	94.71 ± 1.59	97.79 ± 0.26	85.87 ± 4.98	81.37 ± 2.50	85.21 ± 2.42	2.70

The bold values stand for high performance.

TABLE XI
AVERAGE ACCURACY OF SELECTED FEATURES VARYING WITH DIFFERENT H

M\D	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	Average	¹ Rank
100	81.97 ± 3.51	86.77 ± 1.51	72.85 ± 3.93	65.56 ± 3.14	86.28 ± 1.36	95.13 ± 1.64	92.32 ± 0.99	96.69 ± 1.56	73.11 ± 5.21	78.24 ± 3.05	82.89 ± 2.59	4.30
200	82.07 ± 3.02	87.66 ± 1.66	76.14 ± 2.84	75.93 ± 5.24	87.32 ± 2.83	95.19 ± 2.22	94.19 ± 2.33	97.68 ± 0.85	83.98 ± 0.51	83.53 ± 2.77	86.37 ± 2.43	2.10
300	77.88 ± 7.71	87.09 ± 3.60	72.22 ± 2.00	69.26 ± 4.19	87.58 ± 3.22	96.06 ± 1.03	95.78 ± 1.30	97.23 ± 0.52	78.29 ± 4.42	83.92 ± 2.22	84.53 ± 3.02	2.70
400	82.97 ± 6.01	84.09 ± 1.18	72.10 ± 3.61	70.93 ± 2.36	88.41 ± 1.59	95.89 ± 0.70	94.67 ± 3.64	97.42 ± 1.83	84.22 ± 3.28	82.75 ± 1.66	85.35 ± 2.59	2.50
500	79.57 ± 3.75	80.95 ± 1.61	64.25 ± 4.78	74.63 ± 3.93	86.59 ± 3.66	96.03 ± 1.03	93.75 ± 1.66	97.88 ± 0.14	82.49 ± 6.26	81.18 ± 1.11	83.73 ± 2.79	3.40

The bold values stand for high performance.

2) *Analysis on Parameter θ :* To test RF-OS²FS's sensitivity to parameter θ , the parameter sets in this experiment are 0.10, 0.25, 0.50, 0.75, and 1.00. The experimental results are recorded in Table X. The Friedman test is applied to the average accuracy of Table X, and the ρ -values is 1.3877e -06. Therefore, different values affect the RF-OS² FS model. Table X shows that $\theta = 0.50$ has the highest average accuracy and the lowest average rank. The Wilcoxon signed-ranks test also confirms that $\theta = 0.50$ has the best effect. Compared with $\theta = 0.10, 0.25, 0.75$, and 1.00, the ρ -values of $\theta = 0.50$ are 0.0859, 0.0244, 9.7656e -04, and 0.0420, respectively.

3) *Analysis on Parameter H:* In this experiment, H values are set to 100, 200, 300, 400, and 500, and the influence of parameter H on the feature selection of the RF-OS²FS model is analyzed. The average accuracy table of different H values is recorded in Table XI. The ρ -values of the Friedman test in Table XI is 2.5997e -06, which shows that different H values affect the accuracy of feature selection of the RF-OS² FS model, so it is necessary to choose the most suitable H value for the RF-OS²FS model. From these results, the average accuracy does not augment with the increase of H values, and the performance of feature selection with $H = 200$ is the best. The ρ -values of the Wilcoxon signed-ranks test with $H = 200$ and $H = 100, 300, 400$, and 500 are 9.7656e -04, 0.0967, 0.2783, and 0.0098, respectively. Therefore, the H value of the RF-OS² FS model is recommended to be set to 200.

VI. CONCLUSION

Following the core ideal of deep forest and 3WD theory, this article proposes an RF-OS²FS model. The RF-OS²FS model connects the LFA-based estimates and the consequent feature selection recursively extracting the optimal features from sparse streaming features, aiming to enhance the data density and representation ability of LFA-based estimates. The RF-OS²FS model combines OS²FS with 3WD theory to deal with the fuzzy feature selection problem, improving the accuracy of feature selection. It offers a novel perspective on addressing uncertain relationships between labels and completed sparse streaming features, potentially inspiring further research into hybrid methods that combine fuzzy logic with other decision-making frameworks.

Experimental results show that the RF-OS²FS model is superior to other state-of-the-art methods, such as both OSFS and OS²FS.

Although the RF-OS²FS model achieves good performance on sparse streaming features, there is still much improvement that may be done in the future. For example, there are many hyperparameters in the RF-OS²FS model that can be adjusted, and future work may optimize these hyperparameters simultaneously via some evolutionary computation algorithms. To implement symmetry and nonnegativity of RF-OS²FS model's constraints can be helpful in improving its representative learning ability. Besides, we plan to apply RF-OS²FS model to real applications, such as disease diagnosis and prediction, investment analysis, fault diagnosis in industrial systems.

The 3WD theory provides an efficient way to deal with uncertainty by dividing the feature space into three regions: positive, boundary, and negative. This division enables more nuanced decision-making in fuzzy feature selection, where traditional two-way decisions often fail to handle uncertain feature relationships. By explicitly addressing fuzziness, the proposed method enhances the ability of fuzzy systems to handle real-world data with inherent uncertainties. Future studies could explore integrating 3WD with other fuzzy learning algorithms or investigate its applicability to various types of uncertainty, such as probabilistic uncertainty. This could lead to the development of more versatile and powerful fuzzy systems.

REFERENCES

- [1] J. Li et al., "Feature selection: A data perspective," *ACM Comput. Surveys*, vol. 50, no. 6, pp. 94:1–94:45, 2018.
- [2] Y. Koren and R. Bell, "Advances in collaborative filtering," in *Recommender Systems Handbook*. Cham, Switzerland: Springer, 2015, pp. 77–118.
- [3] K. H. Yuan, D. Q. Miao, W. Pedrycz, H. Zhang, and L. Hu, "Multi-granularity data analysis with zentropy uncertainty measure for efficient and robust feature selection," *IEEE Trans. Cybern.*, vol. 55, no. 2, pp. 740–752, Feb. 2025, doi: 10.1109/TCYB.2024.3499952.
- [4] K. H. Yuan, D. Q. Miao, W. Pedrycz, W. P. Ding, and H. Y. Zhang, "Ze-HFS: Zentropy-based uncertainty measure for heterogeneous feature selection and knowledge discovery," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 11, pp. 7326–7339, Nov. 2024.
- [5] K. H. Yuan, D. Q. Miao, Y. Y. Yao, H. Y. Zhang, and X. R. Zhao, "Feature selection using zentropy-based uncertainty measure," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 4, pp. 2246–2260, Apr. 2024.

- [6] X. Wu, K. Yu, W. Ding, H. Wang, and X. Zhu, "Online feature selection with streaming features," *IEEE Trans. Pattern Anal.*, vol. 35, no. 5, pp. 1178–1192, May 2013.
- [7] J. Wang, P. Zhao, S. C. Hoi, and R. Jing, "Online feature selection and its applications," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 3, pp. 698–710, Mar. 2013.
- [8] X. Hu, P. Zhou, P. Li, J. Wang, and X. Wu, "A survey on online feature selection with streaming features," *Front. Comput. Sci.-Chi*, vol. 12, no. 3, pp. 479–493, 2018.
- [9] L. Berti-Equille, H. Harmouch, F. Naumann, N. Novelli, and S. Thirumuruganathan, "Discovery of genuine functional dependencies from relational data with missing values," *Proc. VLDB Endowment*, vol. 11, no. 8, pp. 880–892, 2018.
- [10] Y. Tian, K. Zhang, J. Li, X. Lin, and B. Yang, "LSTM-based traffic flow prediction with missing data," *Neurocomputing*, vol. 318, pp. 297–305, 2018.
- [11] R. Narayanan and Y. Narahari, "A shapley value-based approach to discover influential nodes in social networks," *IEEE Trans. Autom. Sci. Eng.*, vol. 8, no. 1, pp. 130–147, Jan. 2010.
- [12] D. Wu, Y. He, X. Luo, and M. Zhou, "A latent factor analysis-based approach to online sparse streaming features selection," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 52, no. 11, pp. 6744–6758, Nov. 2022.
- [13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.
- [14] D. Wu, X. Luo, Y. He, and M. C. Zhou, "A prediction-sampling-based multilayer-structured latent factor model for accurate representation of high-dimensional and sparse data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 3, pp. 3845–3858, Mar. 2024.
- [15] F. Chen, D. Wu, J. Yang, and Y. He, "Online sparse streaming feature selection with uncertainty," in *Proc. IEEE Int. Conf. Netw., Sens. Control (ICNSC)*, Shanghai, China, 2022, pp. 1–6.
- [16] X. Yang, Y. Li, D. Liu, and T. Li, "Hierarchical fuzzy rough approximations with three-way multigranularity learning," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 9, pp. 3486–3500, Sep. 2022.
- [17] J. Zhan, J. Ye, W. Ding, and P. Liu, "A novel three-way decision model based on utility theory in incomplete fuzzy decision systems," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 7, pp. 2210–2226, Jul. 2022.
- [18] J. Deng, J. Zhan, E. Herrera-Viedma, and F. Herrera, "Regret theory-based three-way decision method on incomplete multiscale decision information systems with interval fuzzy numbers," *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 3, pp. 982–996, Mar. 2023.
- [19] S. Perkins and J. Theiler, "Online feature selection using grafting," in *Proc. 20th Int. Conf. Mach. Learn.*, pp. 592–599, 2003.
- [20] J. Zhou, D. P. Foster, R. A. Stine, and L. H. Ungar, "Streamwise feature selection," *J. Mach. Learn. Res.*, vol. 7, pp. 1861–1885, 2006.
- [21] X. Wu, K. Yu, W. Ding, H. Wang, and X. Zhu, "Online feature selection with streaming features," *IEEE Trans. Pattern Anal.*, vol. 35, no. 5, pp. 1178–1192, May 2013.
- [22] P. Zhou, S. Zhao, Y. T. Yan, and X. D. Wu, "Online scalable streaming feature selection via dynamic decision," *Assoc. Comput. Mach.*, vol. 16, no. 5, pp. 1556–4681, 2022.
- [23] X. Wang and R. Stadler, "Online feature selection for efficient learning in networked systems," *IEEE Trans. Netw. Serv.*, vol. 19, no. 3, pp. 2885–2898, Mar. 2022.
- [24] K. Yu, X. Wu, W. Ding, and J. Pei, "Scalable and accurate online feature selection for Big Data," *ACM Trans. Knowl. Discov. Data*, vol. 11, no. 2, 2016, Art. no. 16.
- [25] P. Zhou, P. Li, S. Zhao, and X. D. Wu, "Feature interaction for streaming feature selection," *IEEE Trans. Neural Netw. Learn.*, vol. 32, no. 10, pp. 4691–4702, Oct. 2021.
- [26] P. Zhou, Y. Zhang, Z. Ling, Y. Yan, S. Zhao, and X. Wu, "Online heterogeneous streaming feature selection without feature type information," *IEEE Trans. Big Data*, vol. 10, no. 4, pp. 470–485, Apr. 2024.
- [27] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning About Data*, vol. 9, Berlin, Germany: Springer, 1991.
- [28] S. Eskandari and M. M. Javidi, "Online streaming feature selection using rough sets," *Int. J. Approx. Reason.*, vol. 69, pp. 35–57, 2016.
- [29] P. Zhou, X. Hu, P. Li, and X. Wu, "Online feature selection for high-dimensional class-imbalanced data," *Knowl.-Based Syst.*, vol. 136, pp. 187–199, 2017.
- [30] P. Zhou, X. Hu, P. Li, and X. Wu, "Online streaming feature selection using adapted neighborhood rough set," *Inf. Sci.*, vol. 481, pp. 258–279, 2019.
- [31] P. Zhou, X. Hu, P. Li, and X. Wu, "OFS-density: A novel online streaming feature selection method," *Pattern Recognit.*, vol. 86, pp. 48–61, 2019.
- [32] S. Li, K. Zhang, Y. Li, S. Wang, and S. Zhang, "Online streaming feature selection based on neighborhood rough set," *Appl. Soft Comput.*, vol. 113, pp. 273–287, 2021.
- [33] P. Zhou, P. Li, S. Zhao, and Y. Zhang, "Online early terminated streaming feature selection based on rough set theory," *Appl. Soft Comput.*, vol. 113, 2021, Art. no. 107993.
- [34] P. Zhou, N. Wang, and S. Zhao, "Online group streaming feature selection considering feature interaction," *Knowl.-Based Syst.*, vol. 226, 2021, Art. no. 107157.
- [35] C. Luo et al., "RHDOFS: A distributed online algorithm towards scalable streaming feature selection," *IEEE Trans. Parallel Distrib. Syst.*, vol. 34, no. 6, pp. 1830–1847, Jun. 2023.
- [36] D. Wu, X. Luo, M. Shang, Y. He, G. Wang, and M. Zhou, "A deep latent factor model for high-dimensional and sparse matrices in recommender systems," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 51, no. 7, pp. 4285–4296, Jul. 2021.
- [37] D. Wu, M. Shang, X. Luo, and Z. Wang, "An L1-and-L2-norm-oriented latent factor model for recommender systems," *IEEE Trans. Neural Netw. Learn.*, vol. 33, no. 10, pp. 5775–5788, Oct. 2022.
- [38] T. T. He, Y.-S. Ong, and L. Bai, "Learning conjoint attentions for graph neural nets," in *Advances in Neural Information Processing Systems*, A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, Eds. Berlin, Germany: Springer, 2021.
- [39] T. T. He, L. Bai, and Y.-S. Ong, "Vicinal vertex allocation for matrix factorization in networks," *IEEE Trans. Cybernet.*, vol. 52, no. 8, pp. 8047–8060, Aug. 2022.
- [40] T. T. He, Y. Liu, T. H. Ko, K. C. C. Chan, and Y.-S. Ong, "Contextual correlation preserving multiview featured graph clustering," *IEEE Trans. Cybernet.*, vol. 50, no. 10, pp. 4318–4331, Oct. 2020.
- [41] Z. Z. Zhou and J. Feng, "Deep forest," *Nat. Sci. Rev.*, vol. 6, no. 1, pp. 74–86, 2019.
- [42] M. Pang, K. M. Ting, P. Zhao, and Z.-H. Zhou, "Improving deep forest by screening," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 9, pp. 4298–4312, Sep. 2022.
- [43] Z. Chen, T. Wang, H. Cai, S. K. Mondal, and J. P. Sahoo, "BLB-gcForest: A high-performance distributed deep forest with adaptive subforest splitting," *IEEE Trans. Parallel Distrib. Syst.*, vol. 33, no. 11, pp. 3141–3152, Nov. 2022.
- [44] W. Zheng, L. Yan, C. Gou, and F. Y. Wang, "Fuzzy deep forest with deep contours feature for leaf cultivar classification," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 12, pp. 5431–5444, Dec. 2022.
- [45] Y. Yuan, Q. He, X. Luo, and M. Shang, "A multilayered-and-randomized latent factor model for high-dimensional and sparse matrices," *IEEE Trans. Big Data*, vol. 8, no. 3, pp. 784–794, Mar. 2022.
- [46] Y. Y. Yao, "Three-way decisions with probabilistic rough sets," *Inf. Sci.*, vol. 180, pp. 341–353, 2010.
- [47] D. D. Guo, W. H. Xu, Y. H. Qian, and W. P. Ding, "Fuzzy-granular concept-cognitive learning via three-way decision: Performance evaluation on dynamic knowledge discovery," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 2, pp. 1409–1423, Feb. 2023.
- [48] D. D. Guo, C. M. Jiang, R. X. Sheng, and S. S. Liu, "A novel outcome evaluation model of three-way decision: A change viewpoint," *Inf. Sci.*, vol. 607, pp. 1089–1110, 2022.
- [49] C. Gao, J. Zhou, J. M. Xing, and X. D. Yue, "Parameterized maximum-entropy-based three-way approximate attribute reduction," *Int. J. Approx. Reason.*, vol. 151, pp. 85–100, 2022.
- [50] K. Y. Liu, T. R. Li, X. B. Yan, H. R. Ju, X. Yang, and D. Liu, "Feature selection in threes: Neighborhood relevancy, redundancy, and granularity interactivity," *Appl. Soft Comput.*, vol. 146, 2023, Art. no. 110679.
- [51] Z. H. Qi, H. Li, K. Zhang, and J. H. Dai, "An attribute fuzzy concept-oriented three-way utility decision model in multi-attribute environments," *Appl. Soft Comput.*, vol. 143, 2023, Art. no. 110353.
- [52] C. Fu, K. Y. Qi, K. Pang, J. Wu, and E. Zhao, "BTWM-HF: A behavioral three-way multi-attribute decision-making method with hesitant fuzzy information," *Expert Syst. Appl.*, vol. 249, 2024, Art. no. 123733.
- [53] T. Y. Yin et al., "Feature selection for multilabel classification with missing labels via multi-scale fusion fuzzy uncertainty measures," *Pattern Recognit.*, vol. 154, 2024, Art. no. 110580.
- [54] T. Y. Yin, H. M. Chen, Z. Yuan, J. H. Wan, K. Y. Liu, and Shi-Jinn Horng, "A robust multilabel feature selection approach based on graph structure considering fuzzy dependency and feature interaction," *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 12, pp. 4516–4528, Dec. 2023.
- [55] D. Liu, "The effectiveness of three-way classification with interpretable perspective," *Inf. Sci.*, vol. 567, pp. 237–255, 2021.

- [56] T. Tlili and S. Krichen, "A simulated annealing-based recommender system for solving the tourist trip design problem," *Expert Syst. Appl.*, vol. 186, 2021, Art. no. 115723.
- [57] W. H. Xu, D. D. Guo, Y. H. Qian, and W. P. Ding, "Two-way concept-cognitive learning method: A fuzzy-based progressive learning," *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 6, pp. 1885–1899, Jun. 2023.
- [58] W. H. Xu, D. D. Guo, J. S. Mi, Y. H. Qian, K. Y. Zheng, and W. P. Ding, "Two-way concept-cognitive learning via concept movement viewpoint," *IEEE Trans. Neural. Netw. Learn. Syst.*, vol. 34, no. 10, pp. 6798–6812, Oct. 2023.
- [59] K. Yu, W. Ding, and X. D. Wu, "LOFS: Library of online streaming feature selection," *Knowl.-Based Syst.*, vol. 113, pp. 1–3, 2016.



Ruiyang Xu (Graduate Student Member, IEEE) received the B.S. degree in information and computing science from the Qufu Normal University, Qufu, China, in 2020 and the Electronic Information degree in computer science from the School of Computer Science and Information Engineering, Harbin Normal University, Harbin, China, in 2022. She is currently working toward the joint Ph.D. degree in computer science with CQUPT united training by the Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing, China.

Her research interests include streaming feature selection, big data analysis, and algorithm design for large-scale data applications.



Di Wu (Member, IEEE) received the Ph.D. degree in computer science from the Chongqing Institute of Green and Intelligent Technology (CIGIT), Chinese Academy of Sciences (CAS), Beijing, China, in 2019.

After his Ph.D., he joined CIGIT, CAS, China, and is currently a Professor of the College of Computer and Information Science, Southwest University, Chongqing, China. He has more than 70 publications, including 24 IEEE Transactions papers and several conference papers on AAAI, IEEE International Conference on Data Mining, World Wide Web, *International Journal of Computing and Artificial Intelligence*, etc. His research interests include machine learning and data mining.

Dr. Wu is currently an Associate Editor for the *Neurocomputing* and *Frontiers in Neurorobotics*.



Xin Luo (Fellow, IEEE) received the B.S. degree in computer science from the University of Electronic Science and Technology of China, Chengdu, China, in 2005 and the Ph.D. degree in computer science from Beihang University, Beijing, China, in 2011.

He is currently a Professor of Data Science and Computational Intelligence with the College of Computer and Information Science, Southwest University, Chongqing, China. He has authored or coauthored more than 400 papers (including more than 160 IEEE Transactions/Journal papers) in the areas of *Artificial Intelligence* and *Data Science*.

Dr. Luo was the recipient of the Outstanding Associate Editor Award from IEEE ACCESS in 2018, IEEE/CAA JOURNAL OF AUTOMATICA SINICA in 2020, and from IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS in 2022–2024. He is a Fellow of AAIA. He is currently an Associate Editor for IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, and IEEE/CAA JOURNAL OF AUTOMATICA SINICA.